

```
In [196]: import numpy as np
import pandas as pd

data = pd.read_fwf("auto-mpg.data")
```

```
In [197]: data.columns=['mpg', 'cylinders', 'displacement', 'horsepower', 'weight', 'acceleration', 'model year', 'origin', 'car name']
```

```
In [198]: data.shape
```

```
Out[198]: (397, 9)
```

```
In [199]: print(data.head(5))
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year \
0	15.0	8	350.0	165.0	3693.0	11.5	70
1	18.0	8	318.0	150.0	3436.0	11.0	70
2	16.0	8	304.0	150.0	3433.0	12.0	70
3	17.0	8	302.0	140.0	3449.0	10.5	70
4	15.0	8	429.0	198.0	4341.0	10.0	70

	origin	car name
0	1	"buick skylark 320"
1	1	"plymouth satellite"
2	1	"amc rebel sst"
3	1	"ford torino"
4	1	"ford galaxie 500"

```
In [200]: data.columns
```

```
Out[200]: Index(['mpg', 'cylinders', 'displacement', 'horsepower', 'weight', 'acceleration', 'model year', 'origin', 'car name'], dtype='object')
```

```
In [201]: print(data.info)
```

```
<bound method DataFrame.info of
mpg  cylinders  displacement  horsepower
weight  acceleration  \
0      15.0          8      350.0      165.0   3693.0          11.5
1      18.0          8      318.0      150.0   3436.0          11.0
2      16.0          8      304.0      150.0   3433.0          12.0
3      17.0          8      302.0      140.0   3449.0          10.5
4      15.0          8      429.0      198.0   4341.0          10.0
..      ...          ...          ...          ...          ...
392    27.0          4      140.0      86.00   2790.0          15.6
393    44.0          4       97.0      52.00   2130.0          24.6
394    32.0          4      135.0      84.00   2295.0          11.6
395    28.0          4      120.0      79.00   2625.0          18.6
396    31.0          4      119.0      82.00   2720.0          19.4

      model year  origin      car name
0           70      1  "buick skylark 320"
1           70      1  "plymouth satellite"
2           70      1    "amc rebel sst"
3           70      1    "ford torino"
4           70      1  "ford galaxie 500"
..          ...      ...          ...
392          82      1  "ford mustang gl"
393          82      2    "vw pickup"
394          82      1  "dodge rampage"
395          82      1  "ford ranger"
396          82      1  "chevy s-10"

[397 rows x 9 columns]>
```

```
In [202]: data.mean()
```

```
/var/folders/f6/clcxz2jd0dd7fkbg9gtzj63h0000gn/T/ipykernel_12649/53190338
6.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions
(with 'numeric_only=None') is deprecated; in a future version this will
raise TypeError. Select only valid columns before calling the reduction.
```

```
data.mean()
```

```
Out[202]: mpg          23.528463
cylinders          5.448363
displacement      193.139798
weight          2969.080605
acceleration      15.577078
model year        76.025189
origin            1.574307
dtype: float64
```

```
In [203]: data.dtypes
```

```
Out[203]: mpg                float64
cylinders                int64
displacement            float64
horsepower              object
weight                  float64
acceleration            float64
model year              int64
origin                  int64
car name                object
dtype: object
```

```
In [204]: data = data.replace({'?': 'NaN'})
data.head()
```

```
Out[204]:
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin	car name
0	15.0	8	350.0	165.0	3693.0	11.5	70	1	"buick skylark 320"
1	18.0	8	318.0	150.0	3436.0	11.0	70	1	"plymouth satellite"
2	16.0	8	304.0	150.0	3433.0	12.0	70	1	"amc rebel sst"
3	17.0	8	302.0	140.0	3449.0	10.5	70	1	"ford torino"
4	15.0	8	429.0	198.0	4341.0	10.0	70	1	"ford galaxie 500"

```
In [205]: data.iloc[:, :8] = data.iloc[:, :8].astype('float')
# data.dtypes
print(data.head(5))
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	\
0	15.0	8.0	350.0	165.0	3693.0	11.5	
1	18.0	8.0	318.0	150.0	3436.0	11.0	
2	16.0	8.0	304.0	150.0	3433.0	12.0	
3	17.0	8.0	302.0	140.0	3449.0	10.5	
4	15.0	8.0	429.0	198.0	4341.0	10.0	

	model	year	origin	car name
0		70.0	1.0	"buick skylark 320"
1		70.0	1.0	"plymouth satellite"
2		70.0	1.0	"amc rebel sst"
3		70.0	1.0	"ford torino"
4		70.0	1.0	"ford galaxie 500"

```
In [206]: # data = pd.read_fwf("auto-mpg.data", na_values='?')
# data.dtypes
```

```
In [207]: data.describe()
```

```
Out[207]:
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	
count	397.000000	397.000000	397.000000	391.000000	397.000000	397.000000	397.000000	3
mean	23.528463	5.448363	193.139798	104.404092	2969.080605	15.577078	76.025189	
std	7.820926	1.698329	104.244898	38.518732	847.485218	2.755326	3.689922	
min	9.000000	3.000000	68.000000	46.000000	1613.000000	8.000000	70.000000	
25%	17.500000	4.000000	104.000000	75.000000	2223.000000	13.900000	73.000000	
50%	23.000000	4.000000	146.000000	93.000000	2800.000000	15.500000	76.000000	
75%	29.000000	8.000000	262.000000	125.000000	3609.000000	17.200000	79.000000	
max	46.600000	8.000000	455.000000	230.000000	5140.000000	24.800000	82.000000	

```
In [216]: data.groupby(by='mpg').describe().horsepower
```

```
Out[216]:
```

	count	mean	std	min	25%	50%	75%	max
mpg								
9.0	1.0	193.00	NaN	193.0	193.00	193.0	193.00	193.0
10.0	2.0	207.50	10.606602	200.0	203.75	207.5	211.25	215.0
11.0	4.0	187.00	28.213472	150.0	172.50	194.0	208.50	210.0
12.0	6.0	185.00	23.528706	160.0	170.25	180.0	193.50	225.0
13.0	20.0	158.35	21.534000	129.0	145.00	152.5	171.25	215.0
...
43.4	1.0	48.00	NaN	48.0	48.00	48.0	48.00	48.0
44.0	1.0	52.00	NaN	52.0	52.00	52.0	52.00	52.0
44.3	1.0	48.00	NaN	48.0	48.00	48.0	48.00	48.0
44.6	1.0	67.00	NaN	67.0	67.00	67.0	67.00	67.0
46.6	1.0	65.00	NaN	65.0	65.00	65.0	65.00	65.0

129 rows × 8 columns

```
In [218]: data.isnull()
```

```
Out[218]:
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin	car name
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...
392	False	False	False	False	False	False	False	False	False
393	False	False	False	False	False	False	False	False	False
394	False	False	False	False	False	False	False	False	False
395	False	False	False	False	False	False	False	False	False
396	False	False	False	False	False	False	False	False	False

397 rows × 9 columns

```
In [219]: data.isnull().sum()
```

```
Out[219]: mpg                0
cylinders                  0
displacement              0
horsepower                6
weight                   0
acceleration              0
model year                0
origin                   0
car name                 0
dtype: int64
```

```
In [220]: data.info
```

```
Out[220]: <bound method DataFrame.info of      mpg  cylinders  displacement  horse
power  weight  acceleration  \
0      15.0        8.0        350.0        165.0   3693.0          11.5
1      18.0        8.0        318.0        150.0   3436.0          11.0
2      16.0        8.0        304.0        150.0   3433.0          12.0
3      17.0        8.0        302.0        140.0   3449.0          10.5
4      15.0        8.0        429.0        198.0   4341.0          10.0
..      ...      ...      ...      ...      ...      ...
392    27.0        4.0        140.0        86.0   2790.0          15.6
393    44.0        4.0         97.0        52.0   2130.0          24.6
394    32.0        4.0        135.0        84.0   2295.0          11.6
395    28.0        4.0        120.0        79.0   2625.0          18.6
396    31.0        4.0        119.0        82.0   2720.0          19.4

      model year  origin      car name
0          70.0    1.0  "buick skylark 320"
1          70.0    1.0  "plymouth satellite"
2          70.0    1.0    "amc rebel sst"
3          70.0    1.0    "ford torino"
4          70.0    1.0  "ford galaxie 500"
..      ...      ...      ...
392         82.0    1.0  "ford mustang gl"
393         82.0    2.0    "vw pickup"
394         82.0    1.0  "dodge rampage"
395         82.0    1.0  "ford ranger"
396         82.0    1.0  "chevy s-10"

[397 rows x 9 columns]>
```

```
In [221]: data.fillna(data.min()).describe()
```

```
Out[221]:
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	
count	397.000000	397.000000	397.000000	397.000000	397.000000	397.000000	397.000000	3
mean	23.528463	5.448363	193.139798	103.521411	2969.080605	15.577078	76.025189	
std	7.820926	1.698329	104.244898	38.885908	847.485218	2.755326	3.689922	
min	9.000000	3.000000	68.000000	46.000000	1613.000000	8.000000	70.000000	
25%	17.500000	4.000000	104.000000	75.000000	2223.000000	13.900000	73.000000	
50%	23.000000	4.000000	146.000000	92.000000	2800.000000	15.500000	76.000000	
75%	29.000000	8.000000	262.000000	125.000000	3609.000000	17.200000	79.000000	
max	46.600000	8.000000	455.000000	230.000000	5140.000000	24.800000	82.000000	

```
In [222]: data['mpg'].value_counts()
```

```
Out[222]: 13.0    20
          14.0    19
          15.0    16
          18.0    16
          26.0    14
          ..
          31.9     1
          16.9     1
          18.2     1
          22.3     1
          44.0     1
          Name: mpg, Length: 129, dtype: int64
```

```
In [ ]:
```